

# SEMI-BLIND SOURCE SEPARATION VIA SPARSE REPRESENTATIONS AND ONLINE DICTIONARY LEARNING

Sirisha Rambhatla and Jarvis Haupt

University of Minnesota-Twin Cities  
Department of Electrical and Computer Engineering  
Minneapolis, MN 55455

## ABSTRACT

This work examines a semi-blind source separation problem where the aim is to separate one source, whose local (nominally periodic) structure is partially or approximately known, from another *a priori* unspecified but structured source, given only a single linear combination of the two sources. We propose a novel separation technique based on local sparse approximations; a key feature of our procedure is the *online* learning of dictionaries (using only the data itself) which sparsely model the *a priori* unknown source. We demonstrate the performance of our proposed approach via simulation in a stylized audio source separation problem.

**Index Terms**— Semi-blind source separation, sparse representations, online dictionary learning

## 1. INTRODUCTION

The blind source separation (BSS) problem entails separating a collection of signals, each comprised of a superposition of some unknown sources, into their constituent components. A canonical example of the BSS task arises in the so-called *cocktail party problem*, and a number of methods have been proposed to address this problem. Perhaps the most well-known among these is independent component analysis (ICA) [1], where the sources are assumed to be independent non-Gaussian random vectors. Other approaches entail more classical matrix factorization techniques like principal component analysis (PCA) [2–4], or, when appropriate for the underlying model, non-negative matrix factorization (NNMF) [5].

Here we focus on a slightly different, and often more challenging setting – the so-called *single channel* source separation problem – where only a single mixture of the source signals is observed. Single channel source separation problems require the use of some additional *a priori* knowledge about the sources and their structure in order to perform separation [6–9]. Here, we assume that the local structure of one of the source signals is *approximately* known (in a manner described in more detail below), and our aim is to separate this partially known source from an unknown “background” source. Our task is motivated by an audio processing application in law enforcement scenarios where electroshock devices are used. A key forensic task in these scenarios is to determine, from audio data recorded by the device itself, the resistive load encountered by the device (corresponding to qualitatively “low” and “high” resistance loads). The approach proposed here can aim to separate the audio corresponding to a nominally periodic and approximately known (up to the resistive load ambiguities) discharge from otherwise unknown, but often highly structured, background audio. The separated audio

signal can subsequently be used to classify the state of the resistive load (we consider only the separation task here).

Our separation approach is based on local sparse approximations of the mixture data. A novel feature of our proposed method is in our representation of the unknown background source – we describe a technique for learning (from the data itself) a *model* that sparsely represents the unknown background source, using tools from the dictionary learning literature (see, eg., [10–12]). The next section describes the problem we consider here more formally, and discusses the nature of our contributions in the context of existing works in sparse representation, dictionary learning, and low-rank modeling.

## 2. BACKGROUND AND PROBLEM FORMULATION

Our effort here is motivated by a single-channel semi-blind audio source separation problem, in which the goal is to separate a nominally periodic and approximately known signal from unknown but structured background interference, given only a superposition of the two sources. Let  $x \in \mathbb{R}^n$  represent our observed data, and suppose that  $x$  may be decomposed as a sum of two sources – one of which ( $x_p \in \mathbb{R}^n$ ) exhibits local structure that is partially or approximately known, and the other ( $x_u \in \mathbb{R}^n$ ) is unknown. In our motivating audio application for example,  $x$  is comprised of samples of an underlying continuous time waveform, and we consider  $x_p$  to be samples of a source that is a nominally regular repetition of one of a small number of prototype signals. One example scenario where this model is applicable is the case where  $x_p$  is, up to some unknown offset jitter, periodic. Our aim is to separate the sources  $x_p$  and  $x_u$  from observations of  $x$ , which may be noisy or otherwise corrupted.

Our proposed approach is based on the principle of local sparse approximations. In order to state our overall problem in generality, we describe an equivalent model for our data  $x$  that facilitates the local analysis inherent to our approach. Let us suppose that  $m$  is an integer that divides  $n$  evenly, such that  $n/m = q$ , an integer. Then  $x \in \mathbb{R}^n$  may be represented equivalently as a  $m \times q$  matrix  $X$ :

$$X = X_p + X_u, \quad (1)$$

where  $X_p$  is a matrix whose columns are non-overlapping length- $m$  segments of  $x_p$ , and similarly for  $X_u$ . The goal of our effort is, in essence, to separate  $X$  into its constituent matrices  $X_p$  and  $X_u$ .

As alluded above, our separation approach entails leveraging local structure in each of the components of  $X$ . Our main contribution comes in the form of a procedure that, given our “partial” information about the columns of  $X_p$ , enables us to learn in an online fashion and from the data itself a dictionary  $D$  such that columns of  $X_u$  are accurately expressed as linear combinations of (a small number of)

Email: {rambh002, jdhaupt}@umn.edu.

columns of  $D$ . In a broader sense, our work is related to some classical approximation approaches as well as several recent works on matrix decomposition. We briefly describe these background and related efforts here, in an effort to put our main contribution in context.

## 2.1. Prior Art

### 2.1.1. Low Rank and Robust Low Rank Approximation

Consider the model (1) and suppose that the columns of  $X_p$  can each be represented as a linear combination of some  $r$  linearly independent vectors, implying that  $X_p$  is a matrix of rank  $r$ . Now, different separation techniques may be employed depending on our assumptions of  $X_u$ . Perhaps the simplest case is where  $X_u$  is random noise (e.g., having entries that are iid zero-mean Gaussian); in this case, the problem amounts to a denoising problem, which can be solved using ideas from low-rank matrix approximation. In particular, it is well-known that the approximation  $\hat{X}_p$  obtained via the truncated (to rank  $r$ ) singular value decomposition (SVD) of  $X$  is a solution of the optimization

$$\hat{X}_p = \arg \min_{L, \text{rk}(L) \leq r} \|X - L\|_F^2, \quad (2)$$

where  $\text{rk}(L)$  is the function that returns the rank of  $L$ .

It is well-known that certain (non-Gaussian) forms of interference  $X_u$  may cause the accuracy of estimators of the low-rank component obtained via truncated SVD to degrade significantly. This is the case, for example, when  $X_u$  is comprised of sparse large (in amplitude) impulsive noise. In these cases, the low-rank approximation problem can be modified to its *robust* counterpart, which goes by the name of robust PCA in the literature [13, 14]. The robust PCA approach aims to simultaneously estimate both the low-rank  $X_p$  and the sparse  $X_u$ , by solving the convex optimization

$$\{\hat{X}_p, \hat{X}_u\} = \arg \min_{L, S} \|L\|_* + \lambda \|S\|_1 \quad \text{subject to } X = L + S, \quad (3)$$

where  $\lambda > 0$  is a regularization parameter. Here  $\|L\|_*$  denotes the *nuclear norm* of  $L$ , which is the sum of the singular values of  $L$ . The nuclear norm is a convex relaxation of the non-convex rank function  $\text{rk}(L)$ . Further,  $\|S\|_1$  is the sum of the absolute entries of  $S$  – essentially the  $\ell_1$  norm of a vectorized version of  $S$ , which is a convex relaxation of the non-convex  $\ell_0$  quasinnorm that counts the number of nonzeros of  $S$ .

Here, of course, we explicitly assume that  $X_u$  is more highly structured, making the separation problem more well-suited to a new suite of techniques that explicitly exploit such structure.

### 2.1.2. Low Rank Plus Sparse in a Known Dictionary

A useful extension of the robust PCA approach arises in the case where  $X_u$  is not itself sparse, but possesses a sparse representation in some known dictionary or basis. One example is the case where the background source is locally smooth, implying it can be sparsely represented using a few low-frequency discrete cosine transform or Fourier basis elements. Formally, suppose that for some known matrix  $D$ , we have that  $X_u = DA_u$ , where the columns of  $A_u$  are sparse. The components of  $X$  can be estimated by solving the following optimization [15]

$$\{\hat{X}_p, \hat{A}_u\} = \arg \min_{L, A} \|L\|_* + \lambda \|A\|_1 \quad \text{subject to } X = L + DA \quad (4)$$

Note that an estimate  $\hat{X}_u$  of  $X_u$  may be obtained directly as  $\hat{X}_u = D\hat{A}_u$ . This approach assumes (implicitly) a priori knowledge of a

dictionary that sparsely represents the background signal, which may be a restrictive assumption in practice.

### 2.1.3. Morphological Component Analysis

A more general model arises when  $X_p$  is not low-rank, but instead, its columns are also sparsely represented in a known dictionary. Suppose that  $X_p$  and  $X_u$  are sparsely represented in some known dictionaries  $D_1$  and  $D_2$ , such that  $X_p = D_1A_1$  and  $X_u = D_2A_2$ , and that the columns of  $A_1$  and  $A_2$  are sparse. Such models were employed in recent work on Morphological Component Analysis (MCA) [16–18], which aimed to separate a signal into its component sources based on structural differences codified in the columns of the known dictionaries. The MCA decomposition can be accomplished by solving the following optimization

$$\begin{aligned} \{\hat{A}_1, \hat{A}_2\} &= \arg \min_{A_1, A_2} \|X - D_1A_1 - D_2A_2\|_F^2 \\ &\text{subject to } \|A_1\|_1 + \|A_2\|_1 \leq \lambda, \end{aligned} \quad (5)$$

for some  $\lambda > 0$ , where the estimates of  $X_p$  and  $X_u$  are formed as  $\hat{X}_p = D_1\hat{A}_1$  and  $\hat{X}_u = D_2\hat{A}_2$ , respectively. When  $X_p$  and  $X_u$  are each comprised of a single column, this optimization is equivalent to the so-called Basis Pursuit (or more specifically, Basis Pursuit Denoising) technique [19], which formed a foundation of much of the recent work in sparse approximation. Note that, as with the previously mentioned approach, this approach also assumes a priori knowledge of a dictionary that sparsely represents the background.

## 2.2. Our Contribution: “Semi-blind” Morphological Component Analysis

Our focus here is similar to the MCA approach above, but we assume only one of the dictionaries, say  $D_1$ , is known. In this case, the MCA approach transforms into a semi-blind separation problem where we try to also learn a dictionary  $D_2$  to represent the unknown signal. Our main contribution comes in the form of a “Semi-Blind” MCA procedure, designed to solve the following modified form of the MCA decomposition

$$\begin{aligned} \{\hat{A}_1, \hat{A}_2, \hat{D}_2\} &= \arg \min_{A_1, A_2, D_2} \|X - D_1A_1 - D_2A_2\|_F^2 \\ &\text{subject to } \|A_1\|_1 + \|A_2\|_1 \leq \lambda, \end{aligned} \quad (6)$$

and this problem forms the basis of the remainder of this paper. Specifically, in Section 3 we propose a procedure, based on alternating minimization, for obtaining local solutions to optimizations of the form (6). In Section 4 we examine the performance of our proposed approach in an application motivated by an audio source separation problem in audio forensics. Finally, we discuss conclusions and possible extensions in Section 5.

## 3. SEMI-BLIND MCA

As described above, our model assumes that the data matrix  $X$  can be expressed as the superposition of two component matrices,  $X_p$  and  $X_u$ . Further, we assume that each of the component matrices possesses a sparse representation in some dictionary, such that  $X_p \approx D_1A_1$  and  $X_u \approx D_2A_2$ , where  $D_1$  is known a priori. Our essential aim, then, is to identify an estimate  $\hat{A}_1$  of the coefficient matrix  $A_1$  and estimates  $\hat{D}_2$  and  $\hat{A}_2$  of the matrices  $D_2$  and  $A_2$ . Our estimates of the separated components are then given by  $\hat{X}_p = D_1\hat{A}_1$ , and  $\hat{X}_u = \hat{D}_2\hat{A}_2$ .

---

**Algorithm 1** Semi-Blind MCA Algorithm

---

**Input:** Original Data  $X \in \mathbb{R}^{m \times q}$ , Known Dictionary  $D_1 \in \mathbb{R}^{m \times d}$ ,  
Regularization parameters  $\lambda_1, \lambda_2, \lambda_3 > 0$ ,  
Number of elements in unknown dictionary  $\ell$ .

**Initialize:**  $\tilde{A}_1 \leftarrow \arg \min_{A_1} \|X - D_1 A_1\|_F^2 + \lambda_1 \|A_1\|_1$

(or other suitable initialization depending on the problem.)

**Iterate (repeat until convergence):**

**repeat**

*Dictionary Learning:*

$$\{\tilde{D}_2, \tilde{A}_2\} \leftarrow \arg \min_{D_2, A_2} \|X - D_1 \tilde{A}_1 - D_2 A_2\|_F^2 + \lambda_2 \|A_2\|_1$$

*Coefficient Update:*

$$\begin{aligned} \tilde{D} &= [D_1 \ \tilde{D}_2] \\ [\tilde{A}_1^T \ \tilde{A}_2^T]^T &\triangleq \tilde{A} \leftarrow \arg \min_A \|X - \tilde{D} A\|_F^2 + \lambda_3 \|A\|_1 \end{aligned}$$

**until** convergence

**Output:** Learned dictionary  $\hat{D}_2 \leftarrow \tilde{D}_2$ ,  
Coefficient estimates  $\hat{A}_1 = \tilde{A}_1$ ,  $\hat{A}_2 = \tilde{A}_2$ .

---

We propose an approach to solve (6) that is based on alternating minimization, and is summarized here as Algorithm 1. Let  $\lambda_1, \lambda_2, \lambda_3 > 0$  be user specified regularization parameters. Our initial estimate of coefficients  $A_1$ , corresponding to the coefficients of  $X_p$  in the known dictionary  $D_1$ , is obtained via

$$\tilde{A}_1 = \arg \min_{A_1} \|X - D_1 A_1\|_F^2 + \lambda_1 \|A_1\|_1, \quad (7)$$

which is a simple LASSO-type problem. We then proceed in an iterative fashion, as outlined in the following subsections, for a few iterations or until some appropriate convergence criteria is satisfied. It should be noted that the lack of joint convexity makes the SBMCA algorithm sensitive to initialization. Therefore, any suitable initialization using sparse approximation techniques, depending upon the problem setting, can be employed. This is well illustrated in Section 4, where we consider an audio forensics application.

### 3.1. Dictionary learning stage

Given the estimate  $\tilde{A}_1$ , we can essentially “subtract” the current estimate of  $X_p$  from  $X$ , and apply a dictionary learning step to identify estimates of the unknown dictionary  $D_2$  and the corresponding coefficients  $A_2$ . In other words, we solve

$$\{\tilde{D}_2, \tilde{A}_2\} = \arg \min_{D_2, A_2} \|X - D_1 \tilde{A}_1 - D_2 A_2\|_F^2 + \lambda_2 \|A_2\|_1. \quad (8)$$

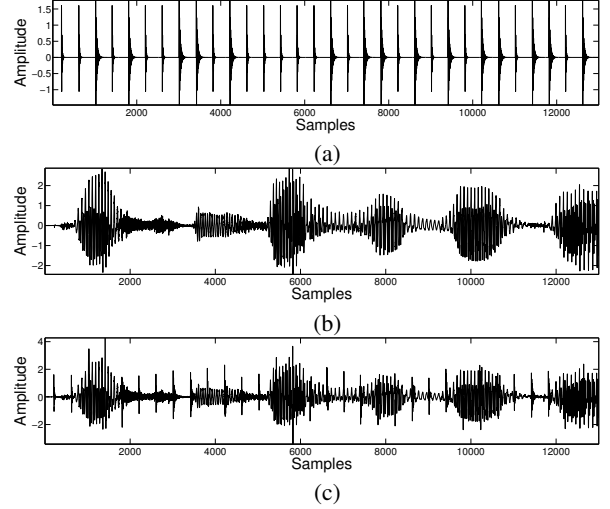
Now, given the estimate  $\tilde{D}_2$ , we update our current estimate of the *overall* dictionary  $\tilde{D} = [D_1 \ \tilde{D}_2]$ . We then update the *overall* coefficient matrix by solving another sparse approximation problem, as described next.

### 3.2. Sparse approximation stage

Given our current estimate of the overall dictionary, we update the corresponding coefficient matrices by solving the following LASSO-like problem:

$$[\tilde{A}_1^T \ \tilde{A}_2^T]^T \triangleq \tilde{A} = \arg \min_A \|X - \tilde{D} A\|_F^2 + \lambda_3 \|A\|_1. \quad (9)$$

Now, we extract the submatrix  $\tilde{A}_1$  from  $\tilde{A}$ , and repeat the overall processing (beginning with the dictionary learning step). These steps are iterated until some appropriate convergence criteria is satisfied.



**Fig. 1:** A segment of mixture components (noise free): (a) the nominally periodic signal  $x_p$  (each segment is the discharge corresponding to one of the two resistive load states, randomly selected); (b) the background signal  $x_u$ ; (c) the mixture  $x$ .

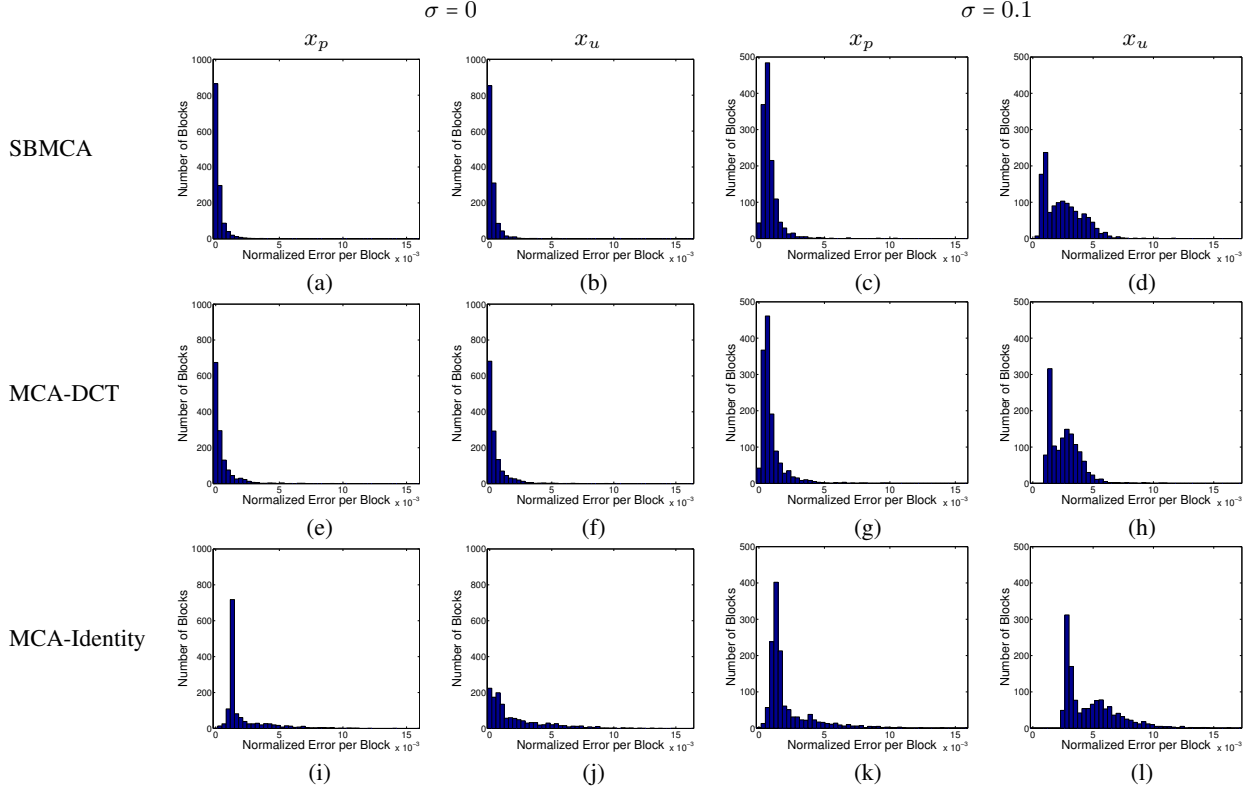
## 4. EVALUATION: AN APPLICATION IN AUDIO FORENSICS

We demonstrate the performance of our approach on a stylized version of the audio separation task described in the introduction, which is motivated by forensic examination of audio obtained during law enforcement events where electroshock devices are utilized. For the sake of this example, we suppose that the electroshock devices discharge approximately 36 times per second, and the waveforms generated by the device during discharge take one of two different forms depending on the level of resistive load encountered by the device. The collected audio corresponds to the nominally periodic discharge of the device, superimposed with background noise (e.g., speech). Our aim is to separate this superposition into its components.

Figure 1 shows a segment of the signals used in the simulation. We simulate the form of the approximately periodic signals ( $x_p$ ), shown in Figure 1 (a), using two distinct exponentially decaying sinusoids, to emulate a series RLC circuits with different parameters, to model the loaded and open circuit states. Specifically, we generate two distinct waveforms, which correspond to the two states (high and low resistive load), and form the overall signal  $x_p$  by concatenating randomly-selected versions of these prototype signals, each of which is subject to a few samples of timing offset in order to model the non-idealities of the actual electroshock device. A speech signal<sup>1</sup> shown in Figure 1 (b), was used to model background noise that may be present during the altercation. We simulate the overall raw audio data as a linear combination of  $x_p$ ,  $x_u$  and zero-mean random Gaussian noise  $\mathcal{N}(0, \sigma^2)$  (Figure 1 (c) depicts the ideal case  $\sigma = 0$ ).

The data matrix  $X$  is formed from the signal  $x$  as discussed in Section 2 using non-overlapping segments with 400 samples each, and we form the dictionary  $D_1$  by incorporating certain circular shifts of the nominal prototype pulses from which  $x_p$  was generated. We then employ the semi-blind MCA approach (discussed in Section 3) to separate the background audio from the approximately

<sup>1</sup>Speech Samples obtained from VoxForge Speech Corpus: [www.voxforge.org/home](http://www.voxforge.org/home)



**Fig. 2:** Histogram of normalized error-per-block measured using the vector  $l_2$ -norm of extracted nominally periodic signal  $x_p$  and extracted speech signal  $x_u$  for Semi-blind MCA, and MCA-DCT, and MCA-Identity, for the audio forensic application.

known periodic portion.

We compare the performance of our approach with two versions of MCA, one using the DCT basis and the other using the identity basis to form the dictionary  $D_2$ . We use the estimated  $x_p$ , obtained via MCA-DCT procedure to initialize our approach, as follow: we apply one step of orthogonal matching pursuit (OMP) [20] on the estimate of  $x_p$  obtained via MCA-DCT to form the initial (one component per column) estimate  $A_1$  for the SBMCA algorithm.

Table 1 lists the *best achievable* reconstruction SNRs (in dB) of each method. We note that our interest here is in comparing the best performances achieved by MCA and our proposed method, so we *clairvoyantly* tune the value(s) of the regularization parameter to give the lowest error for each task. (In general, a different regularization parameter may have been utilized to obtain the reconstruction SNRs of each signal component, even for the same method and same noise level – in other words, the SNRs listed may not be jointly achievable from a *single* implementation of any of the stated procedures).

A second, perhaps more interesting, performance comparison is shown Figure 2, which depicts the histogram of normalized errors-per-block, measured using the vector  $l_2$ -norm, for each method<sup>1</sup>. We observe from the distribution of  $l_2$ -errors across blocks, that

<sup>1</sup>Panels (a), (e) and (i) represent the histogram of normalized error-per-block for  $x_p$  and (b), (f) and (j) represent the histogram of normalized error-per-block for  $x_u$  via SBMCA, MCA-DCT and MCA-Identity respectively, with standard deviation of gaussian noise  $\sigma = 0$ . Panels (c), (g) and (k) represent the histogram of normalized error-per-block for  $x_p$  and (d), (h) and (l) represent the histogram of normalized error-per-block for  $x_u$  via SBMCA, MCA-DCT and MCA-Identity respectively, with standard deviation of gaussian noise  $\sigma = 0.1$ .

**Table 1:** Comparative analysis of Reconstruction SNR(in dB).

Noise $\mathcal{N}(0, \sigma^2)$	$\sigma = 0$		$\sigma = 0.1$	
Method \ Signal	$x_p$	$x_u$	$x_p$	$x_u$
<b>SBMCA</b>	<b>23.72</b>	<b>29.32</b>	<b>19.72</b>	<b>16.84</b>
<b>MCA-DCT</b>	20.44	26.02	18.09	16.72
<b>MCA-Identity</b>	10.90	16.06	10.78	11.44

the SBMCA procedure (Figure 2 (a-d)) results in larger number of blocks with lower errors as compared to the MCA-DCT (Figure 2 (e-h)) and MCA-Identity (Figure 2 (i-l)). This feature is of primary importance in the audio forensics application where classifying each period of the nominally periodic signal  $x_p$ , as one of the two prototype signals, is of interest.

## 5. CONCLUSION

We proposed a semi-blind source separation technique based on local sparse approximations. Our approach exploits partial prior knowledge of one of the sources, in the form of a dictionary which sparsely represents local segments of one of the sources. A key feature of our approach is the online learning of a dictionary (from the mixed source data itself) for representing the unknown background source. We posed the problem as an optimization task, proposed a solution approach based on alternating minimization, and verified its effectiveness via simulation in a stylized audio forensics application. Possible extensions to other applications (eg., image and video processing) are left to future efforts.

## 6. REFERENCES

- [1] C. Jutten and J. Herault, "Blind Separation of Sources, Part I: An Adaptive Algorithm Based on Neuromimetic Architecture," *Signal Processing*, vol. 24, no. 1, pp. 1–10, July 1991.
- [2] H. Hotelling, "Analysis of a Complex of Statistical Variables into Principal Components," *Journal of Educational Psychology*, vol. 24, pp. 417–441, 1933.
- [3] C. Eckart and G. Young, "The Approximation of One Matrix by Another of Lower Rank," *Psychometrika*, vol. 1, pp. 211–218, 1936.
- [4] I.T. Jolliffe, *Principal Component Analysis*, Springer Verlag, 1986.
- [5] A. Cichocki, R. Zdunek, and S. Amari, "New Algorithms for Non-Negative Matrix Factorization in Applications to Blind Source Separation," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 5, pp. V, May 2006.
- [6] G. J. Jang and T. W. Lee, "A Maximum Likelihood Approach to Single-Channel Source Separation," *Journal of Machine Learning Research*, vol. 4, pp. 1365–1392, Dec. 2003.
- [7] T. P. Jung, S. Makeig, C. Humphries, T. W. Lee, M. J. McKeown, V. Iragui, and T. J. Sejnowski, "Removing Electroencephalographic Artifacts by Blind Source Separation," *Psychophysiology*, vol. 37, pp. 163–178, 2000.
- [8] M. N. Schmidt and R. K. Olsson, "Single-channel Speech Separation using Sparse Non-negative Matrix Factorization," *International Conference on Spoken Language Processing (INTERSPEECH)*, 2006.
- [9] M.E. Davies and C.J. James, "Source Separation using Single Channel ICA," *Signal Processing*, vol. 87, no. 8, pp. 1819 – 1832, 2007.
- [10] B. A. Olshausen and D. J. Field, "Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1?," *Vision Research*, vol. 37, no. 23, pp. 3311–3325, 1997.
- [11] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: Design of Dictionaries for Sparse Representation," *In Proceedings of SPARS05*, pp. 9–12, 2005.
- [12] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online Learning for Matrix Factorization and Sparse Coding," *Journal of Machine Learning Research*, vol. 11, pp. 19–60, 2010.
- [13] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust Principal Component Analysis?," *Journal of the ACM*, vol. 58, no. 3, pp. 11, 2011.
- [14] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, "Rank-Sparsity Incoherence for Matrix Decomposition..," *Society for Industrial and Applied Mathematics (SIAM) Journal on Optimization*, vol. 21, no. 2, pp. 572–596, 2011.
- [15] M. Mardani, G. Mateos, and G. B. Giannakis, "Recovery of Low-Rank Plus Compressed Sparse Matrices with Application to Unveiling Traffic Anomalies," *IEEE Transactions on Information Theory*, 2012, Online: [arXiv:1204.6537v1](https://arxiv.org/abs/1204.6537v1) [cs.IT].
- [16] J. L. Starck, M. Elad, and D. L. Donoho, "Image Decomposition via the Combination of Sparse Representations and a Variational Approach," *IEEE Transactions on Image Processing*, vol. 14, no. 10, pp. 1570–1582, 2005.
- [17] D. L. Donoho and G. Kutyniok, "Microlocal Analysis of the Geometric Separation Problem," *CoRR*, vol. abs/1004.3006, 2010.
- [18] J. Bobin, J. L. Starck, J. Fadili, Y. Moudden, and D. L. Donoho, "Morphological Component Analysis: An Adaptive Thresholding Strategy," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2675–2681, 2007.
- [19] S. Chen and D. Donoho, "Basis Pursuit," in *Conference Record of the Twenty-Eighth Asilomar Conference on Signals, Systems and Computers*. IEEE, 1994, vol. 1, pp. 41–44.
- [20] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal Matching Pursuit: Recursive Function Approximation with Applications to Wavelet Decomposition," in *Proceedings of the 27th Annual Asilomar Conference on Signals, Systems and Computers*, 1993, pp. 40–44.